Diabetes, Blood pressure, and Heart disease Prediction using Web Application

Prof. M. GOUTHAMAN	GIRISHA M V[1]	PITCHMA PRIYA K [2]	SHRUTHI M [3]
Department of Computer Science	Department of Computer Science	Department of Computer Science	Department of Computer Science
Jeppiaar Engineering College	Jeppiaar Engineering College	Jeppiaar Engineering College	Jeppiaar Engineering College
Chennai, India	Chennai, India	Chennai, India	Chennai, India

I. INTRODUCTION

Diabetes, high blood pressure, and heart diseases are among the most common diseases that affect millions of people worldwide. These diseases are not only costly to manage, but they also have severe long-term consequences on an individual's health and well-being. Early detection of these diseases is crucial for the successful management and prevention of complications.

Machine learning has shown great potential in predicting and diagnosing various diseases. In this project, we aim to develop a web application that predicts the risk of diabetes, high blood pressure, and heart disease using machine learning algorithms. The application will provide users with an easy-to-use interface to input their health data and receive a prediction of their disease risk.

II. ABSTRACT

Diabetes, high blood pressure, and heart disease are widespread chronic diseases that affect millions of people worldwide. Early detection and

management of these diseases are crucial to prevent complications and improve outcomes. Machine learning has shown promise in predicting and diagnosing various diseases, and this project aims to leverage this technology to develop a web application that predicts the risk of diabetes, high blood pressure, and high cholesterol.

The application will provide users with an intuitive interface to input their health data, such as age, gender, body mass index (BMI), blood pressure, fasting glucose, and lipid profile. The machine learning models used in the application will be trained on a large data set of health records and will use a variety of features to predict the risk of each disease.

The application will provide users with a clear and easy-to-understand prediction of their disease risk and will also provide recommendations for lifestyle changes that may help reduce their risk. The application will be tested on a sample population to

validate its accuracy and effectiveness in predicting disease risk.

This project aims to develop a web application that leverages machine learning enabling early detection and prevention of these chronic diseases.

The project aims to develop a web application that predicts the risk of three common diabetes, high blood pressure, and heart disease. The application will use machine learning algorithms to analyze health data provided by the user, such as age, gender, body mass index (BMI), blood pressure, fasting glucose, and lipid profile. Based on these inputs, the application will predict the likelihood of the user developing one or more of these chronic diseases.

To achieve this goal, the project will involve several stages. First, we will gather a large dataset of health records that includes information on patients' demographics, medical history, and laboratory test results. This dataset will be used to train machine learning models that can predict the risk of diabetes, high blood pressure, and heart disease based on a variety of features.

Next, we will develop a web application that provides users with an easy-to-use interface to input their health data and receive a prediction of their disease risk. The application will use the trained machine learning models to analyze the user's data and provide a clear and easy-to-understand prediction of their risk. The application will also provide users with recommendations to predict the risk of diabetes, high blood pressure, and heart disease. The application will provide users with an easy-to-use interface to input their health data and receive a prediction of their disease risk,
for lifestyle changes that may help reduce their risk.

To ensure the accuracy and effectiveness of the application, we will test it on a sample population and evaluate its performance in predicting disease risk. We will also collect user feedback to identify areas for improvement and make necessary adjustments to the application.

The end goal of the project is to develop a web application that can help individuals assess their risk of developing chronic diseases and take proactive steps to prevent or manage them. By leveraging machine learning technology, we aim to provide users with a powerful tool to improve their health outcomes and overall quality of life.

Data Collection: The project will involve collecting a large dataset of health records from multiple sources, such as hospitals, clinics, and research studies. The dataset will be carefully curated to ensure that it includes a diverse range of patients with varying health conditions, demographics, and lifestyles. We will also take steps to ensure the privacy and security of the data, following relevant regulations and best practices.

Feature Selection: Once we have the dataset, we will conduct an extensive analysis to identify the most relevant features for predicting the risk of diabetes,

high blood pressure, and heart disease This may involve using statistical techniques, such as correlation analysis and feature importance ranking, to determine which features are most strongly associated with each disease.

Machine Learning Models: We will train several machine learning models to predict the risk of each disease, using a variety of algorithms such as logistic regression, decision trees, and neural networks. We will evaluate the performance of each model using standard metrics such as accuracy, precision, recall, and F1 score, and select the best-performing model for each disease.

Web Application Development: The web application will be developed using modern web technologies such as HTML, CSS, and JavaScript, along with a server-side programming language such as Python or Node.js. We will use a popular web application framework such as Flask or Express.js to streamline development and ensure scalability and maintainability. The application will be designed to be user-friendly and accessible, with clear and concise instructions for inputting data and interpreting results.

Testing and Evaluation: Once the application is developed, we will conduct rigorous testing to ensure that it is accurate, reliable, and robust. We will test the application on a sample population and compare its predictions to actual disease outcomes to evaluate its performance. We will also collect user feedback to identify areas for improvement and make necessary adjustments to the application.

Overall, the project will involve a comprehensive and systematic approach to developing a web application that leverages machine learning to predict the risk of diabetes, high blood pressure, and high cholesterol. The application has the potential to be a valuable tool for individuals seeking to improve their health outcomes and reduce their risk of chronic disease.

III.HELPFUL HINTS

In this medical field, it is essential to predict and diagnose diseases to prevent them early. Blood Pressure, Diabetes, and Heart disease are a few of the serious conditions arising due to metabolism inadequacy. To predict and diagnose these diseases Machine Learning (ML)algorithms are valuable for disease prediction. This system based on predictive modeling predicts the disease of the user based on the symptoms that the user provides as input to the system. The system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the random forest classifier and also, Implement Blood Pressure, Diabetes, and Heart disease using Machine Learning and in the proposed method it gives the better accuracy.

Causes of Diabetes

Genetic factors are the main cause of diabetes. It is caused by at least two mutant genes in chromosome 6.

Causes of Blood Pressure

The more you weigh the more blood flow you need to supply oxygen and nutrients to your tissue much salt (sodium) in your diet. Too little potassium in your diet.Lack of

is physically active. Drinking too much alcohol.

Causes of Heart Disease

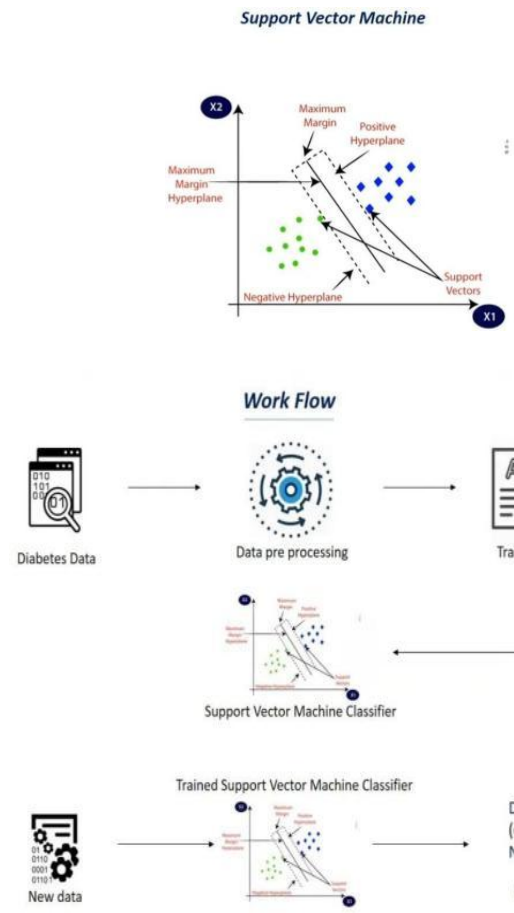
The reasons of a particular form of heart disease vary. Heart disease comes in a variety of forms.The following are some conditions or common causes of arrhythmias

Cardiomyopathy, Coronary artery disease, Diabetes, Drug misuse, Emotional stress, Excessive use of alcohol or caffeine, Heart problems present at birth (congenital heart defects)High blood pressure

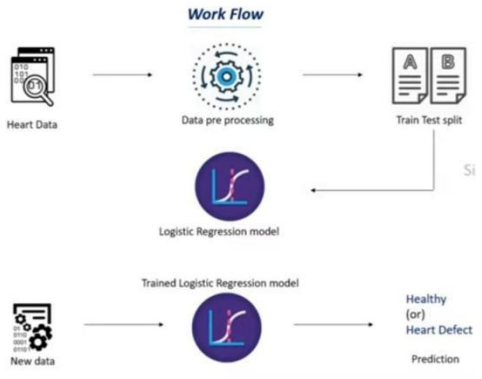
We have mixed structured and unstructured data in the healthcare fields to determine disease risk in this project.

The use of a latent factor model to recreate missing data in medical records obtained from online sources. We could also assess the major chronic diseases in a specific area and population using statistical information. We consult hospital experts to learn about useful features when dealing with structured data. We apply the random forest technique to automatically identify characteristics from unstructured text files..

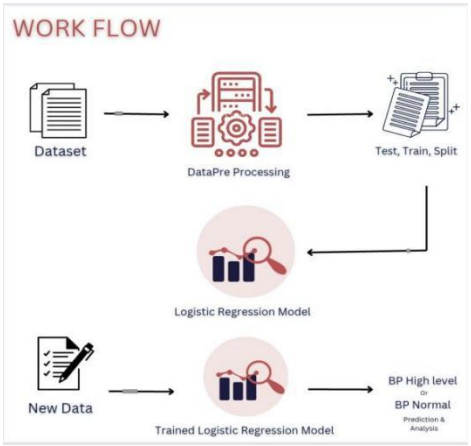
DIABETES WORKFLOW



HEART DISEASE WORKFLOW



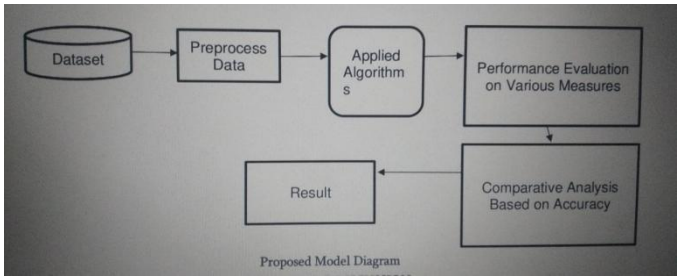
BLOOD PRESSURE WORKFLOW



Building Model

Numerous styles are used to perform data mining. Machine literacy is one of the approaches. Since bracket ways are used in this design, bracket is one of the data mining processes in this phase of categorical data bracket. And this step is divided into two phases training and testing. In the training phase, destined data and associated class markers are used for bracket. The training stage is frequently appertained to as supervised literacy. The medication and testing phases of the bracket process are depicted in the illustration. In the training process, training tuples are used, and in the test data phase, test data tuples are used, and the bracket rule's delicacy is calculated. Assume that the bracket rule's delicacy on testing data is sufficient for the rule to be used for the bracket of unlined data. This emphasis is on the class imbalance problem and the need to handle this problem before applying any algorithm to achieve better delicacy rates. The class imbalance substantially occurs in a dataset having dichotomous values, which means that the class variable has two possible issues and can be handled fluently if observed before in the data preprocessing stage and will help in boosting the delicacy of the prophetic model.

PROPOSED MODEL FOR DIABETES, BLOOD PRESSURE, AND HEART



Complaint Machine literacy is asub-field of Artificial Intelligence. Machine literacy algorithms help computers to make opinions without demanding unequivocal coding. These algorithms fed literal data and also make prognostications about unborn events. This makes them briskly and more important than other styles that bear hand-rendering rules. For illustration, the recommendation system is a common use case. Other common uses include fraud discovery, spam, malware pitfalls, business processes, and prophetic conservation.

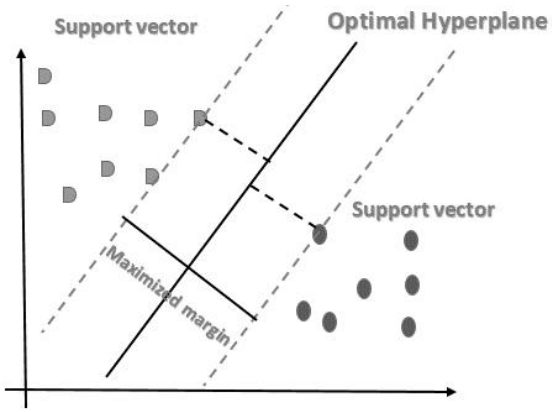
LOGISTIC REGRESSION

Retrogression is a fashion that aims to model the relationship between several characteristics and a nonstop target variable. Retrogression is a machine learning fashion in which a model predicts labors that are nonstop figures. sphere areas where retrogression is generally used are finance, investment, soothsaying, time series

modeling and chancing the cause- and-effect relationship between variables, the relationship between gadarene driving and the number of road accidents can be better anatomized using retrogression, etc

SUPPORT VECTOR MACHINE

Support Vector Machine is one of the most popular literacy algorithms in machine literacy. SVM can be used for both bracket and retrogression. Primarily SVM is used for working bracket problems. SVM is generally used to deal with small and complex types of datasets. SVM follows a unique approach compared to other machine learning algorithms. The working of SVM is as follows SVM creates an n- dimensional space called a hyperplane, which segregates the classes present in the dataset and the hyperplanes also help descry being data points. The decision to produce a hyperplane depends on the size of the inputfeatures.However, If the input features are assumed to betwo.the hyperplane will be treated as a single line and if there are more features than two, the hyperplane will be treated as a two- dimensional aeroplane



. Now the focus is on optimized hyperplane hunt that can insulate classes well. This can be decided using the periphery gap. The data points closest to the hyperplanes are called support vectors. The larger the periphery gap between the support vectors, the lesser the chance of choosing an optimized hyperplane.

SVM are of two types

1) Linear SVM Classifier When the data is classified into only two classes using a straight- line hyperplane, also the classifier is called a direct SVM.

2) Non-Linear Classifier When the data isn't classified into two classes, the classifier is called a nonlinear SVM Logistic retrogression Logistic Retrogression is one of the most common bracket algorithms.

Logistic regression:
Logistic Regression is one of the most common classification algorithms.

DIABETES MODULE

Training accuracy	
Testing accuracy	

BLOOD PRESSURE MODULE

Training accuracy	
Testing accuracy	

HEART DISEASE MODULE

Training accuracy	
Testing accuracy	

IV.UNITS

Diabetes is mostly measured based on the sugar levels and solving present in the human blood. The sugar-contained liquid is consumed, which contains glucose, and blood will be drawn to check blood sugar levels. A normal result is 140 mg/dL or lower. If your level is higher than 140 mg/dL, the procedure finalized is to take a glucose tolerance test. Units used are measured in terms of 140 milligrams per decilitre. Blood Pressure is measured based on an electronic device which is calculated in terms of a millimeter of mercury (mmHg). 120/80mmHg is the normal pressure of the blood

Heart Disease i.e cardiovascular diseases is measured based on the levels of BP(mmHg), Cholesterol(mg/dL), BMI Index value, oxygen saturation levels, and Weight(Kg) of a particular mentioned patient. Further shows that heart disease is the final call for the above-mentioned factors.

V. SOME COMMON MISTAKES

Diagnostic errors are most common, the above modal is solemnly based on the patient's accurate inputs for prediction and diagnostic methods. Wrong diagnosis, leading to severe cases or recommendation of improper medical specifications.

Most revised models chose prediction due to lack of accuracy and well-developed architectures, furthermore, the diagnostic part of any disease has been ignored due to reckoning for medical approval or misleading measurements of medicine consumed by the patient.

Predictions are concluded and predicted for widespread diseases, not for the uncommon. Even though prediction plays a vital role though some of them may have no diagnosis or prevent the disease by offering medications.

Thus concluded models and further models in the future are introduced to find and conduct prediction, and diagnosis rates for 100 percent.

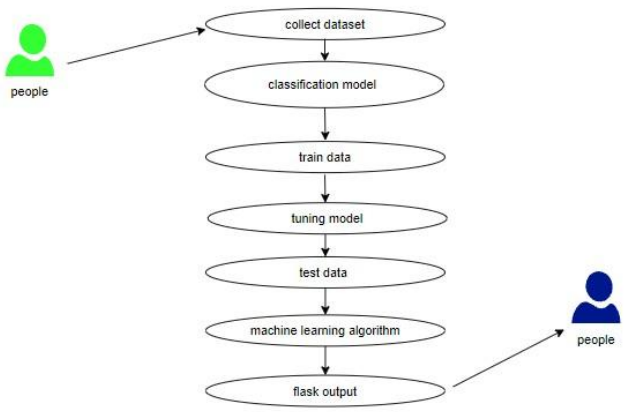
VI. ACKNOWLEDGEMENT

The system is introduced and evolved due to the mentorship of

Dr. J. Anitha Gnanaselvi, the assistant professor in the Department of Computer Science and Engineering. Along with the guidance of Mr. M. Goudhaman, a professor in the Department of Computer Science and Engineering, at Jeppiaar Engineering College.

Our humble consolidation to the students, professors, and scientists referenced below for the development of web application systems on Diabetes, Blood pressure, and Heart disease Prediction.

VII.USE CASE DIAGRAM



VIII. REFERENCES

[1] F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, Computer Vision and Machine Intelligence in Medical Image Analysis. London, U.K.: Springer, 2019.

[2] World Health Organization (WHO). (2020). WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019. Accessed: Oct. 22, 2021. [Online]. Available: https://www.who.int/news/item/09-12-2020-whoreveals-leading-causes-of-death-and-disability-worldwide-2000-2019

[3] A. Frank and A. Asuncion. (2010). UCI Machine Learning Repository. Accessed: Oct. 22, 2021. [Online]. Available: http://archive.ics.uci.edu/ml

[4] G. Pradhan, R. Pradhan, and B. Khandelwal, "A study on various machine learning algorithms used for prediction of diabetes mellitus," in Soft Computing Techniques and Applications (Advances in Intelligent Systems and Computing), vol. 1248. London, U.K.: Springer, 2021, pp. 553–561, doi 10.1007/978-981-15-7394-1_50.

[5] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using a soft voting classifier," Int. J. Cogn. Comput. Eng., vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.

[6] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, ‘‘Prediction of diabetes using machine learning algorithms in healthcare,’’ in Proc. 24th Int. Conf. Autom. Comput. (ICAC), Sep. 2018, pp. 6–7, doi: 10.23919/IConAC.2018.8748992.

[7] S. K. Dey, A. Hossain, and M. M. Rahman, ‘‘Implementation of a web application to predict diabetes disease: An approach using machine learning algorithm,’’ in Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT), Dec. 2018, pp. 21–23, doi: 10.1109/ICCITECHN.2018.8631968.

[8] A. Mir and S. N. Dhage, ‘‘Diabetes disease prediction using machine learning on big data of healthcare,’’ in Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA), Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697439.

[9] S. Saru and S. Subashree. Analysis and Prediction of Diabetes Using Machine Learning. Accessed: Oct. 22, 2022. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368308

[10] P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC), Mar. 2019, pp. 367–371, doi 10.1109/ICCMC.2019.8819841.

[11] S. Wei, X. Zhao, and C. Miao, ‘‘A comprehensive exploration to the machine learning techniques for diabetes identification,’’ in Proc. IEEE 4th World Forum Internet Things (WF-IoT), Feb. 2018, pp. 291–295, doi: 10.1109/WF-IoT.2018.8355130.

[12] M. F. Faruque and I. H. Sarker, ‘‘Performance analysis of machine learning techniques to predict diabetes mellitus,’’ in Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE), Feb. 2019, pp. 7–9, doi: 10.1109/ECACE.2019.8679365.

[13] B. Jain, N. Ranawat, P. Chittora, P. Chakrabarti, and S. Poddar, ‘‘A machine learning perspective: To analyze diabetes,’’ Mater. Today: Proc., pp. 1–5, Feb. 2021, doi: 10.1016/J.MATPR.2020.12.445.

[14] N. B. Padmavathi, ‘‘Comparative study of kernel SVM and ANN classifiers for brain neoplasm classification,’’ in Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT), Jul. 2017, pp. 469–473, doi: 10.1109/ICICICT1.2017.8342608.

[15] J. Liu, J. Feng, and X. Gao, ‘‘Fault diagnosis of rod pumping wells based on support vector machine optimized by improved chicken swarm optimization,’’ IEEE Access, vol. 7, pp. 171598–171608, 2019, doi: 10.1109/ACCESS.2019.2956221.

[16] P. Chinas, I. Lopez, J. A. Vazquez, R. Osorio, and G. Lefranc, ‘‘SVM and ANN application to multivariate pattern recognition using scatter data,’’ IEEE Latin Amer. Trans., vol. 13, no. 5, pp. 1633–1639, May 2015, doi: 10.1109/TLA.2015.7112025.

[17] Y. Yang, J. Wang, and Y. Yang, ‘‘Improving SVM classifier with prior knowledge in microcalcification detection1,’’ in Proc. 19th IEEE Int. Conf. Image Process., Sep. 2012, pp. 2837–2840, doi: 10.1109/ICIP.2012.6467490.

[18] J. Liu, J. Feng, Q. Xiao, S. Liu, F. Yang, and S. Lu, "Fault diagnosis of rod pump oil well based on support vector machine using preprocessed indicator diagram," in Proc. IEEE 10th Data-Driven Control Learn. Syst. Conf. (DDCLS), May 2021, pp. 120–126, doi: 10.1109/DDCLS52934.2021.9455702.

[19] P. Patil, N. Yaligar, and S. M. Meena, "Comparision of performance of classifiers–SVM, RF, and ANN in potato blight disease detection using leaf images," in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Dec. 2017, pp. 1–5, doi: 10.1109/ICCIC.2017.8524301.